

A Digest of the Main Insights and Achievements of the Project IM-CLeVeR – Intrinsically Motivated Cumulative Learning Versatile Robots

*An EU funded Integrated Project (IP), Grant Agreement number 231722 ,
Duration of project: 1st May 2009 – 30th April 2013*

Project website: www.im-clever.eu

Authors:

*Gianluca Baldassarre, Marco Mirolli, Elisabetta Visalberghi, Eugenio
Guglielmelli, Flavio Keller, Peter Redgrave, Kevin Gurney, Martin
McGuinnity, Jochen Triesch, Mark Lee, Juergen Schmidhuber,
Andrew Barto*

2013

Contact: Gianluca Baldassarre (Coordinator), Consiglio Nazionale delle Ricerche, Istituto di Scienze e
Tecnologie della Cognizione (CNR-ISTC) Phone: +39 06 44595231 E-mail: gianluca.baldassarre@istc.cnr.it

Table of Contents

Introduction.....	2
How do results on IM from the human/animal behavioural experiments constrain and direct future development of neuroscience and psychological hypotheses, or robotic learning and control method	3
How do results on intrinsic motivations from the neuroscience modelling constrain and direct future development of robotic learning and control methods, and the design of future neuroscience and behavioural hypotheses?.....	5
How do results on intrinsic motivations from robotic experiments constrain and direct future research on autonomously learning robots and neuroscience/behavioural hypotheses?.....	8
List of major detailed achievements from single partners.....	11
Where the project converged to a consensus position.....	13
Notes on the multiple distinct approaches that need to be investigated and reconciled in future work	14

Introduction

This document is directed to illustrate the main achievements and insights gained by the IM-CLeVeR project. These achievements suggest new hypothesis of research on intrinsic motivations and autonomous cumulative open-ended learning both for running new neuroscience and psychology experiments and for building future autonomously developing robots.

IM-CLeVeR is a project funded by the European Commission under the 7th Framework Programme (FP7/2007-2013), "Challenge 2 - Cognitive Systems, Interaction, Robotics", grant agreement No. ICT-IP-231722. Much information on the project can be found in its web-site: www.im-clever.eu.

We thank the Project Officer Cécile Huet, and the Project Evaluators, Luc Berthouze, Ben Kuipers, and Yasuo Kuniyoshi, for their guidance and steering that have importantly contributed to the achievements illustrated this document.

How do results on IM from the human/animal behavioural experiments constrain and direct future development of neuroscience and psychological hypotheses, or robotic learning and control method

We illustrate here some key achievements and insights from the empirical experiments of IM-CLeVER. To our opinion, these should inform future empirical research (both neuroscientific and psychological) on intrinsic motivations and also be exploited for building robots capable of a truly open-ended autonomous development.

The development of the Joystick task paradigm for studying intrinsic motivations. Various versions of the joystick task involving action discovery by rodents, monkeys, and humans are now in use by collaborating groups, in Japan, New Zealand, Spain, Germany, France, and Ireland. For the last two decades, experiments implicating the brain's dopamine system in behavioural reinforcement were limited by the use of stimuli that evoked phasic dopamine responses at pre-saccadic, pre-attentive latencies. Collaborative studies with a Japanese group confirmed in monkeys that visual reinforcement that engages both cortical and subcortical sensory processing is more effective than either alone. We have now discovered that sensory input to the brain's basic reinforcement mechanism is not restricted to primitive sub-cortical sensory processing but involve also cortical processes. We are so now confident that more sophisticated events that are of intrinsic interest can reinforce the acquisition of novel behaviour, which, at some later stage can be deployed if the elicited event becomes associated with extrinsic reward. The above findings demonstrate clearly that sophisticated cortical processing can also access the brain's dopamine reinforcement system. This information confirms the range of sensory events that can be used to reinforce intrinsically or extrinsically motivated action acquisition.

Recommendation for models: This removes an important constraint from biomimetic models of reinforcement as we have at least two sources of intrinsic rewards, those generated by sub-cortical structures and those generated by cortical structures.

The components of action. A relevant conceptual development forced by the IM-CLeVER project is the realisation that the acquisition of a novel actions frequently involves the independent learning of WHERE, WHAT, WHEN and HOW the actions have to be performed. It is difficult to imagine that each of these aspects will be acquired in the same class of neural network. This raises the possibility that, in the same way that different aspects of sensory perception are conducted in spatially segregated networks, so the different aspects of action may be processed likewise. In both cases, distributed processing will require a mechanism to 'bind' the different aspects of perception and action into a unitary whole. Our recognition that the 'binding problem' applies to action as well as perception is an important advance for both neuroscience and the development of biomimetic artificial agents.

Recommendation for models: The insight that action can actually be decomposed into sub-components, and that distinct brain areas characterized by distinct functioning and learning processes underlying them, can have important implications for computational models as it suggests a divide-and-conquer strategy that might greatly facilitate the autonomous acquisition of actions in robots, and their flexible re-use.

The importance of agency within intrinsic motivations in children. In children, spontaneous exploration plays a fundamental role in the learning process providing subjects with an increasingly diverse set of opportunities for acquiring, practicing and refining new abilities. The project studied which strategies children adopt to spontaneously learn action outcome and if and how they are able to use these knowledge in a different context. The results show that unexpected and surprising stimuli trigger children exploration even without an extrinsic goal. Such curiosity-triggered exploration seems to be kept alive by the contingency between children's actions and platform's outcomes, especially in older children whose curiosity vanished sooner in the case the contingency was missing. The action-outcome contingency acts as a facilitator increasing the likelihood of experiencing the effects related to the action to learn. The repeated exploration of that particular action or that specific button controlling the observed effect seems to be fundamental for this processes.

Intrinsic motivations based on action-outcome contingencies are strong also in monkeys. The empirical results

on monkeys suggest that, in the absence of extrinsic reinforcement, the opportunity to discover action-outcome contingencies promotes individuals' exploratory drives and learning. From a neuroscientific perspective, the rapid decrease in exploration shown by yoked subjects closely matches what expected at the brain level. In fact, in the absence of behaviourally rewarding consequences, the phasic DA response toward unpredicted novel neutral stimuli diminishes rapidly due to habituation. Under this condition subjects are prevented from further exploring the board. By contrast, when outcomes are contingent with actions (like in the experimental condition) they may function as primary rewards, reinforcing action repetition and thus learning. So far, little is known on how action-outcome contingencies may effectively block the decrease of phasic DA response. This research question could be further addressed by combining the neuroscientific approach (i.e., through non-invasive EEG recording of visual evoked potentials in humans) with behavioural observations on monkeys and children. The study with monkeys illustrates the importance of combining neuroscientific results with behavioural ones, and might possibly help to address future studies in the neuroscientific field.

General implications. Based on these results, an important advancement from our behavioural experiments with monkeys and children has been that the discovery of agency is itself strongly intrinsically motivating. This opens a new area for neuroscience and psychology. Important future experiments in behavioural neuroscience will have to test the specific mechanisms that allow the detection of agency, and how the detection of agency can block the normal process of habituation ("things get boring") that leads to seek other experiences. The contrast would be with the rate of habituation associated with unpredicted sensory events whose onset was uncorrelated with any behavioural output.

Recommendation for models: These insights on agency might also be exploited to build an intrinsic motivation engine capable of driving a truly open-ended cumulative development in robots. For example, it would provide a rationale for including an 'agency-bonus' in reinforcement algorithms.

Extrinsic motivations are very strong in non human primates, and tend to cover intrinsic motivations. As also shown by the theoretical investigations of the project, the relation between intrinsic and extrinsic motivations is very important for organisms. The experiments with monkeys and children are indicating that for (adult) monkeys this relation is even stronger than for humans (and children), in the sense that the extrinsic motivation component is so strong that many efforts had to be spent to let the intrinsic motivations to emerge in the experiments. Children, instead, seem to have a very strong intrinsic motivation component that can be more easily captured in experiments, and studied under many facets.

Recommendation for models: The arbitration between extrinsic and intrinsic motivations, leading to give the control, or to mix, the two, is a key problem to solve in computational models.

Advancements of our understanding of Parkinson. A review of the action-outcome learning in Parkinson patients led to an influential publication (Nature Reviews Neuroscience) but also the abandonment of the originally proposed investigation. This work confirmed the fundamental distinction between the habitual and the goal-directed systems supporting action because only the latter (at least at initial phases of the disease) is strongly impaired in Parkinson patients: their behaviour is only controlled by goal-directed processes.

Recommendation for models: In order to build robots as flexible as organisms it is paramount to build controllers that can implement both habitual and goal-directed like behaviours and learning processes.

Pivotal importance of attentional processes. The empirical experiments strongly indicated the importance that attention plays in learning processes driven by intrinsic motivations. Indeed, a main way in which intrinsic motivations guide learning is by guiding attention towards specific stimuli that deserve learning resources and time.

Recommendation for models: After this important insight, attention has gained a central position, steadily increasing in time, within the CLEVER-B demonstrators (an important part of the biologically constrained robotic models of IM-CLeVeR). This insight should also inform future models on intrinsic motivations.

Pivotal importance of goals. The comparison of the results with monkeys and humans indicated that a major difference that might distinguish the two species might involve their capacity to pivot the behaviour on goals. Indeed, direct observation and (indirectly) data seem to indicate that humans are more strongly driven by goals than monkey when they act. Goals are a powerful means to focus behaviour and learning on single

objects/events, to resist distractors, and to capture key causal links between actions and outcomes (which are potential future goals). Moreover, goals might be a powerful source of learning signals once they are accomplished, as indirectly indicated by the fact that when children succeed in accomplishing a goal they manifest this in various ways, for example by smiling, looking at the caregivers, triggering non-functional gestures, etc.

Recommendation for models: The intuition on the importance of goals has already informed the latest CLEVER-B models. Indeed, a key element in these models has been the acquisition of the link between actions and their outcomes (“goals” when they become desired). These links allow for example a later recall of suitable actions when the outcome becomes desirable (hence a goal). In CLEVER-B3/4 goals were also exploited to implement biological versions of the inhibitors of dopamine initially hardwired in CLEVER-B2. Future models should assign a pivotal role to goals.

How do results on intrinsic motivations from the neuroscience modelling constrain and direct future development of robotic learning and control methods, and the design of future neuroscience and behavioural hypotheses?

Modelling constrained by data from brain and behaviour furnish a number of insights testable in future empirical experiments, and important indications for future robotic models :

The paramount importance of attention to guide learning and behaviour. As also suggested by the empirical experiments, the neuroscientific models of IM-CLeVeR (e.g., those of the CLEVER-B Demonstrators) clearly indicated that attention is paramount for autonomous learning and for exhibiting a flexible behaviour. This appears clear if one realises that attention has to do with the important “where” component of actions. By attention we refer to both the bottom-up mechanisms, that drive sensors to focus on highly-informative regions of space, and to the top-down mechanisms, that drive sensors to focus on regions of space that are important to accomplish the task of the agent. Attention is important for a number of reasons for learning: (a) it drives learning resources on relevant aspects of the environment, e.g. it protects from distractors; (b) if driven by novelty, it can direct exploratory processes towards novel objects. Attention is also important for functioning: (a) it allows the drastic reduction of information to be processed; (a) it greatly supports action selection processes (e.g., “if I look at object A then I do action 1, if I look at object B then I do action B”).

Recommendations for future robotic methods: Controllers guiding autonomous robots should have a component driving learning and functioning of attention as this can completely change the problems faced by the processes driving the learning of pragmatic actions (e.g., reaching/grasping actions). The attention component should encompass both bottom-up and top-down processes.

Recommendations for future psychology/neuroscience experiments: Empirical research tends to study phenomena in isolation as this facilitates their understanding. For example, there is a tension between those studying bottom-up attention processes and top-down attention processes with both tending to say that attention is mainly of one or the other type. Our models instead show that they are both crucial for attention, and that very interesting phenomena can be observed when they interact. Thus, empirical research should devote much studies to investigate the interactions between bottom-up and top-down attention processes during learning.

The importance of the coupling between attention and pragmatic actions (where and what). The neuroscientific models also indicated how a strong coupling between attention and pragmatic action (e.g., reaching/manipulation actions are always directed where attention is directed, or has been recently directed) can give important computational advantages. The reason of this is that the coupling creates the needed, but at the same time flexible, binding between the “where” and “what” aspects of actions. So, for example, if the agent has learned to perform a certain manipulation action on an object it can perform the same action on the same object, or a different object, located in a different position. This: (a) creates the possibility of generalising the action with respect to “where”; (b) creates the possibility of exploring the effects of an acquired action on novel objects.

Recommendations for future robotic methods: Controllers guiding autonomous robots should have a first component driving learning and functioning of attention, and a second component driving learning and functioning of pragmatic actions; they should implement a strong coupling between the two.

Recommendations for future psychology/neuroscience experiments: Empirical research tends to study phenomena in isolation as this facilitates their understanding. However, the tremendous synergies emerging from the coupling of attention and pragmatic action can be understood only if the two are studied in an integrated fashion. New experimental paradigms and tests are needed to this purpose.

Goals are pivotal for cumulative learning: now we need to understand how they can be self-generated in autonomous agents (animals and robots). The bio-constrained modelling clearly indicated that goals are pivotal for autonomous learning for a number of reasons: (a) they can focus attention and learning, e.g. protect from distractors; (b) they represent a “pointer” through which it is possible to recall the actions that lead to their accomplishment, e.g. to recall the actions when their goals are activated; (c) they facilitate the composition of actions at an abstract level, e.g. to form sequences and hierarchies; (d) they can generate learning signals when accomplished, and these signals can drive the learning of actions directed to them. These aspects are all very important for cumulative learning. These considerations point to a crucial problem: how can autonomous agents (animals or robots) form goals autonomously? In particular, how can they form *useful* goals?

Recommendations for future robotic methods: Focus research on methods usable to endow robots with the capacity to self-generated goals. Knowledge-based intrinsic motivations (e.g., those based on prediction errors, or novelty) are an important candidate for doing this.

Recommendations for future psychology/neuroscience experiments: Focus research on understanding how animals autonomously self-generate goals. An important means for doing this would be to systematically study children at play.

The paramount importance of knowledge transfer processes for cumulative learning. An astonishing aspect of autonomous learning in children is the contrast existing between the initial *slow* learning of the first months of life, and the *exponential* eager learning of the following phases. The computational models of the project suggest that the mechanisms underlying this phenomenon might be the capacity of real organisms to transfer knowledge from already acquired skills to the new skills to be learned to solve new problems. Understanding the mechanisms that support the transfer of knowledge is so of paramount importance to understand autonomous development.

Recommendations for future robotic methods: Aim to develop algorithms and architectures capable of transferring knowledge in relation to information on: representations, action policies, value functions, forward models, and inverse models.

Recommendations for future psychology/neuroscience experiments: Design and run experiments directed to investigate the behavioural processes that support transfer in learning, e.g. in children. Also, design neuroscientific experiments directed to investigate the neural processes underlying transfer in learning, e.g. of motor skills in monkeys.

Repetition bias is a key mechanisms for intrinsically motivated learning. Repetition bias, initially proposed in Redgrave and Gurney in 2006, is a temporary increase of the likelihood to produce a certain action following an initial intrinsic reinforcement. This mechanism might play an important role in acquiring novel actions as it allows a self-training focussing of attention and learning resources on specific stimuli and actions, akin the Piagetian “circular reactions”.

Recommendations for future robotic methods: The repetition bias idea is at the heart of several models developed in IM-CLeVeR (e.g., those of CLEVER-B). However, further investigations are needed to fully understanding the mechanisms through which repetition bias can be implemented, and how it can be coupled to the type of knowledge to be acquired.

Recommendations for future psychology/neuroscience experiments: Experimental work showed how the joystick task is highly suited to investigating the existence and nature of repetition bias. This paradigm might so be used to investigate the unclear aspects of repetition bias in both neuroscientific and psychological experiments directed, for example, to understand the specific brain mechanisms underlying (e.g.: Is it based on transient

changes of synapses or on transient motivations?) and their effect on behaviour (How is repetition best exploited for learning?).

Habituation of intrinsic rewards is a second key mechanism of intrinsically motivated learning. Habituation of intrinsic rewards is the idea that the dopaminergic response to novel outcomes reduces as that outcome becomes predictable. This is observed experimentally in animals. If such habituation does not occur, then the agent will continue to focus obsessively on the same activities leading to a suboptimal behaviour.

Recommendations for future robotic methods: It is important to envisage specific ways to have *transient* intrinsic motivations and learning signals. A critical point of this will be to design mechanisms that create a strong coupling between the fading away of the motivations/learning signals and the termination of the learning processes being guided by such motivations/signals. Indeed, efficient motivation signals should not terminate when the agent is still learning, as it would miss learning opportunities, nor it has to terminate much later after the agent has terminated learning, as this would imply a waste of time.

Recommendations for future psychology/neuroscience experiments: We envisage an experiment test with the joystick task, whereby if habituation to the reinforcement signal could be reduced (e.g. by varying the nature of the signal somehow, or by pharmacological manipulation) then subjects would suboptimally persevere to pursue one goal after performance has achieved the steady state.

Two phases of learning and the function of intrinsic motivations. One key aspect of the tests of the intrinsically motivated models built in IM-CLeVeR is the exploitation of two phases, one where the agent acquires knowledge and skills by intrinsic motivations, and a second one where it exploits the knowledge and skills acquired in the first phase to pursue some well-defined extrinsic tasks. This organisation of the text allows to clearly measure the advantages rendered by intrinsic motivations. This is done by comparing the performance in the second phase of agents that undergo the first phase, and can learn with intrinsic motivations, with "controls" agents that undergo the first phase, but operate in a condition where they cannot acquire the knowledge needed in the second phase. This structure of the experimental paradigm is very powerful to understand intrinsic motivations.

Recommendations for future robotic methods: The importance of intrinsic motivations for robot scan be measured by using tests formed by two phases as just explained. This is very important as very often within the robotic community the nature and function of intrinsic motivations is not very clear, so the availability of an operational means to measure them, where intrinsic and extrinsic (i.e., user-task related) motivations are neatly separated, will have an important role in showing their potential for the advancement of developmental robots.

Recommendations for future psychology/neuroscience experiments: Future experimental paradigms on intrinsic motivations might exploit the general idea of the two phases to better isolate the effects of intrinsic motivations vs. other confounding factors.

There exist multiple intrinsic motivations! The neuroscience models lead us to understand that there are several different intrinsic motivation mechanisms rather than one. So we have: novelty-based intrinsic motivations, prediction-based intrinsic motivations, competence-based intrinsic motivations. And for each of these there are different possible sub-mechanisms (algorithms and architectures) that can be exploited. Those models also established clear links between these mechanisms and possible corresponding mechanisms in brain (e.g., prediction-based IM were related to superior colliculus, novelty-based IM were related to hippocampus).

Recommendations for future robotic methods: Aim to understand which mechanisms can be exploited for which function within robotic controllers, as this might constitute a useful tool-box for building autonomously developing robots.

Recommendations for future psychology/neuroscience experiments: The neuroscientific literature relevant for intrinsic motivations often addresses phenomena that are not directly linked to intrinsic motivations (e.g., prediction failures for attention and arousal; novelty for memory formation). It is so important to produce theoretical works that unify those processes under the research agenda of intrinsic motivations with the goal of highlighting their common fundamental principles. This is expected to foster research not only on intrinsic motivations but also within the original fields of study.

How do results on intrinsic motivations from robotic experiments constrain and direct future research on autonomously learning robots and neuroscience/behavioural hypotheses?

We highlight here the major achievements of the project robotic and machine learning research, highlighting their implications for future research in the same fields and also for suggesting hypotheses for neuroscience and psychology:

An integrated architecture for modular behaviour and for protecting the robot. We developed the modular behavioral framework MoBeE (Frank et al., 2012) for the CLEVER-K Demonstrator during the four years of the project. MoBeE is not exclusive applicable for the IM-CLeVeR project as it is designed for any kind of humanoid or other complex robot. However, it depends on the YARP middleware, which is the software platform for the iCub humanoid robot. MoBeE contains essential components to apply learning tasks on a real physical robot: it contains a model of the robot for fast forward kinematic simulation and self-collision detection and a world model which is used for collision detection and as an abstract representation of the environment. Among other things, this allows safe experiments with the real robot as self-collisions or dangerous collisions with the objects in the environment can be prevented when needed. MoBeE is implemented as a filter within the YARP framework. This allows regular YARP modules to communicate with the controlled robot through a transparent safeguard.

Recommendations for future research in robotics: The MoBeE system can support the construction of complex models to control the iCub robot giving also the possibility of avoiding to damage the robot hardware. For this reason it can be used as an important basis to develop intrinsically motivated self-developing robot controllers.

Suggestions for research in neuroscience/psychology: The experience with the robot iCub indicated that a major problem of open-ended learning is the unpredictability of the interactions that the robot might engage with the environment, and this might lead to dangerous effects on its physical integrity. This generates an interesting question for developmental psychology: how can infants develop all their knowledge and skills on the basis of intrinsic motivations while overcoming the safety problem so clearly appearing in our robotic experiments? For example, what is the role of pain in guiding learning, or the role of low muscular torques of children for their safety?

A robust mechanism for object recognition. We developed a vision module which automatically detects interesting location in the visual field, focusing on objects and trains automatically robust representations of the detected objects. Interesting locations of the scene are selected by saliency maps. The objects are detected with a feature detector on the stereo image and the robust tracking of the objects is based on cartesian genetic programming. This approach is based on a genetic search of possible combinations of a number of filters available in OpenCV: when suitably combined, these filters lead to a robust segmentation and recognition of the object. The detected objects can be joined to the object data base and added to the MoBeE world model.

Recommendations for future research in robotics: The visual system developed is very robust and so might be exploited in future robotic systems where a robust recognition of objects is needed. The approach also shows how powerful learning approaches are in comparison to standard artificial vision approaches when they can work at a higher level (e.g., at the level of the combination of filters). This aspect should be further investigated in the future and might suggest new powerful algorithms and architectures.

Suggestions for research in neuroscience/psychology: The suitable recombination of high-level filters revealed very powerful for our visual systems: is it possible that natural visual systems exploit a similar strategy, e.g. at the highest levels of the visual system?

Learning motion solutions in high-dimensional motor spaces. We developed a robot motion control module which learns task-relevant roadmaps in high-dimensional motor spaces based on natural evolution strategies. A task is a short-term manipulation action like ‘reaching’, ‘pushing sideways’, ‘pushing forward’, and so on. The tasks are defined as constraints with simple mathematical equations in the Euclidean work space for specific control points. A control point can be an end-effector of another part of the robot, e.g. the elbow. The learning algorithm finds solutions in the 41 dimensional joint space which are homogeneously distributed in the 6 dimensional euclidean work space.

Recommendations for future research in robotics: The techniques developed here are very powerful and can be exploited as a basis to develop further systems in the future. For example in the last part of the project we have shown how they can be used to support an intrinsically motivated reinforcement learning controller engaged in solving complex reaching and manipulation tasks. This research might be usefully developed in future work.

Suggestions for research in neuroscience/psychology: How do organisms solve the complex motions problems faced with the proposed models? Are there biological correspondents of the algorithms proposed here, e.g. systems that work in the work space and control the system in the joint redundant space on the basis of loose constraints?

Slow Features Analysis for implementing intrinsic motivations. We developed a curiosity-driven autonomous system for learning perceptual invariances and subsequent skills, called Curious Dr. MISFA, that learns from high-dimensional raw image data, generated from the eyes of an exploring iCub robot. Curious Dr. MISFA enables the iCub to continually learn skills – toppling an object leads to grasping the object, which finally leads to pick & place — starting with no knowledge of its environment, except for a compressed joint-space representation, previously learned by natural evolution. Through CD-MISFA, the robot explores with the goal to acquire perceptual invariances from Slow Feature Analysis, incrementally from the video data. To this end, we use our incremental version of Slow Feature Analysis and the version incorporating auto-encoders for advanced non-linear processing.

Recommendations for future research in robotics: Future work might investigate how to use our approaches based on Slow Features Analysis as a core intrinsic motivation engine usable to support cumulative learning in more sophisticated hierarchical systems. The experiments run so far indicate that this approach might be at the basis of a truly open-ended system.

Suggestions for research in neuroscience/psychology: Slow Feature Analysis was initially inspired as a way in which place cells develop in the hippocampus of rats while they navigate in the environment. Is it possible that the same mechanism is used in other part of the brain to intrinsically drive the autonomous formation of complex sensorimotor representations? This might be tested with physiological recording experiments similar to those used to discover place cells.

Intrinsic motivations for learning visual features, vergence, and motion tracking. We have extended the efficient coding hypothesis for learning sensory representations to active perception. To this end we have combined sparse coding approaches with a form of intrinsically motivated reinforcement learning that favors movements of the sense organs that aid in encoding the sensory data more efficiently. We have demonstrated on the iCub robot how this leads to self-calibrating systems for binocular vision and vergence control as well as motion perception and tracking behavior.

Recommendations for future research in robotics: Intrinsic motivations based on reconstruction errors revealed an extremely versatile means to guide the self development of visual features, vergence, and motion tracking. Future research should investigate other possible applications of this seemingly very powerful approach.

Suggestions for research in neuroscience/psychology: The results indicate that the models illustrated above are very effective. The natural questions to investigate in empirical experiments is hence: does the autonomous development of visual features, vergence, and motion tracking in real brains and animals depend on similar mechanisms?

Learning to recognise objects on the basis of intrinsic motivations. We have developed a curiosity-driven vision system that learns to represent and recognize objects in its environment. It utilizes an attention mechanism that drives the system to look those locations in the environment where it estimates the highest learning progress. The system has been demonstrated on the iCub robot.

Recommendations for future research in robotics: The models developed so far revealed quite powerful. Future research might be so directed to evaluate how far this approach scales up in supporting a cumulative learning of visual representations of objects.

Suggestions for research in neuroscience/psychology: The models suggest questions for psychological experiments: do young children explore the visual scene as done by our models driven by intrinsic motivations?

Exploiting novelty detection for developing multiple aspects of robotic controllers. We also focused on the development of novelty detection methods based on biological learning and habituation. In particular, we developed a new approach for segmentation of objects using live streams from cameras, using a 3D approach that is capable of detecting features and recognizing objects. We also built a new novelty detector learner based on the biological non-associative learning form of habituation, validating the models experimentally with physical robots. We extended core novelty detection methods such that we demonstrated a robot system capable of continuously and autonomously exploring, learning and identifying novel objects within its perceptual and search space based on visual processing and habituation. We developed a learning architecture using an expandable bag-of-words for effective cumulative learning of visual perceptions.

Recommendations for future research in robotics: Our experiments show that novelty detection is a powerful mechanism to support various types of learning processes in robots. We expect that novelty detection mechanisms might be exploited for even further functions in future robot applications.

Suggestions for research in neuroscience/psychology: Since novelty detection has revealed so important for the autonomous learning of robots, we expect that it might have a similar important role in organisms. So, novelty detection might be performed not only on the basis of the well-known novelty-detection processes of hippocampus, but also within other areas of brain. An interesting hypothesis to investigate with physiology experiments.

Using evolutionary techniques to develop complex compositional motor repertoires. In the latter years of the project we explored robot interactions with objects and action learning. Action learning methods were further developed incorporating biologically inspired novelty detection for effective exploration and continuous learning. A fuzzy neural network was used to learn and optimise basic affordances through interactions with objects. Methods were developed to automate composition and parameterization of skills, and learning and adaptation of skills based on novel evolutionary algorithms. Overall, these experiments showed again (as done for vision, see above) that evolutionary learning methods working on the composition of higher level chunks (in this case, parameterised motor primitives composed on the basis of an abstract description) are extremely powerful.

Recommendations for future research in robotics: Future work might further investigate how to exploit these evolutionary techniques to solve more challenging tasks requiring sophisticated motor hierarchies.

Suggestions for research in neuroscience/psychology: The idea of solving motor problems with mechanisms composing chunks of motor behaviour seems such a powerful solution that we expect to be also exploited in real brains in animals. Experimental paradigms should be designed to investigate this.

Exploiting ideas from biology to improve robots on the basis of ideas from CLEVER-B models. In the final year of the project we developed a progression from static to dynamic novelty detection and in particular the incorporation and extension of a neuroscience based model of intrinsic motivation into a practical robotic environment. We implemented two extensions of the CLEVER-B models based on intrinsic motivations on a PR2 robot. The first extension of the model involved incorporating a probabilistic biased selection approach (PBS) based on former acquired knowledge. The second extension involved predictive learning over time. For this work we devised an experimental setup for action learning where the robot interacted with balls on a table, with holes as targets, and a limited selection of dummy actions. Vision modules enabled the robot to locate an object and use intrinsic motivations to learn to focus and track it. The PR2 robot was able to learn representations for various objects based on intrinsic motivations. The new integrated approach showed consistent improved behavior and clear benefits.

Recommendations for future research in robotics: This work is a good example that shows how technologically relevant robotic problems can be solved with controllers strongly inspired by bio-constrained models. Future research might aim to further analyze the multiple bio-constrained models developed in IM-CLeVeR to evaluate if other ideas developed in such models can be exploited for solving machine learning problems.

Suggestions for research in neuroscience/psychology: One version of the model showed the importance of using previous experiences to create a prior bias on the use of actions in future conditions. Can the brain have mechanisms that allow it to collect such general-purpose knowledge on the usability of actions in different conditions?

List of major detailed achievements from single partners

ISTC-CNR-LOCEN

- *Management*: The experience of the project showed that integration, a critical aspect of large-sized European projects, can be fostered by: (a) Exploiting the Demonstrators as the main collector of efforts of the project; (b) giving important roles to junior researchers: (d) taking really seriously the idea that those who want to understand the brain and behaviour of animals with computational models, and those who want to build effective robots, can achieve important breakthroughs by understanding in depth the reciprocal problems, methods, and techniques, and by closely collaborating together.
- The Team contributed to define the mechatronic board and the experimental protocols with monkeys and children: this is now a novel unique experimental tool that can be used to investigate intrinsically motivated cumulative learning in monkeys, children and robots.
- The Team played a key role in developing a general theory on intrinsic and extrinsic motivations; in clarifying the distinction between prediction-based, novelty-based, and competence-based intrinsic motivations; in developing several bio-constrained models, and some machine-learning models, of cumulative learning based on these different types of intrinsic motivations.
- The Team highlighted theoretically, and based on models, how the cumulative learning of skills (competence) based on intrinsic motivations needs to pivot on action *goals* and action-outcome representations.
- The Team developed general theories and system-level bio-constrained models of the hierarchical organisation of brain: these theories and models allows capturing a number of behavioural phenomena related to autonomous cumulative learning, goal-directed behaviour, and habits formation and exploitation.
- The Team developed an integrative bio-constrained model of extrinsic motivations (based on amygdala) and of the brain architecture underlying hierarchical sensorimotor behaviour, based on the main striato-cortical loops (limbic, associative, sensorimotor).
- The Team developed new bio-inspired hierarchical-reinforcement learning models that can solve multiple tasks by suitably allocating "expert modules" to them based on the sensorimotor complexity of the tasks. The model works with continuous states and actions in the iCub robot.
- The Team coordinated and played a major role in the design and implementation of CLEVER-B demonstrators involving CNR, AU and USFD. This is the first bio-constrained robotic model that can learn action-outcome associations based on intrinsic motivations, and can recall them based on the internal re-activation of goals by extrinsic motivations. The CLEVER-B architectures represent important milestones in the construction of bio-constrained models of

ISTC-CNR-UCP

- The Team developed a detailed experimental protocol and experimental set-up to investigate intrinsic motivations in monkeys and participated to define the features of the mechatronic board used in the same experiments.
- The Team ran two sets of experiments with monkeys and the mechatronic board, collected and analysed the data and, on this basis, showed how intrinsic motivations can guide learning in monkeys.

ISTC-CNR-Barto

- Prof. Barto and his Team fostered theoretical and modelling ideas on intrinsic motivations within the whole project, with visits and participations to various activities of the project, in particular with CNR.
- The Team also developed state-of-the-art hierarchical reinforcement learning and intrinsic motivation system.

UCBM

- The Team developed a common technological tool, called "the mechatronic board", to study different experimental populations: children, capuchin monkeys and humanoid robots. The tool was delivered in two release: one for children and robots (three replicas), and one for monkeys (one replica).

- The Team supported the data acquisition sessions in both children and monkey experiments.
- The Team developed software tools in Matlab environment to assist UCBM-LDN for data analysis of children experiment.
- The Team recruited 36 children to test in intrinsic motivation experiments. Twelve children were involved in a pilot study carried out during 2010-2011 to refine the experimental protocol and indexes. The other twenty-four children, aged 3 and 4 years, were involved in the final experiment.
- The Team carried out investigations and modelling of development of rhythmic manipulation skills together with CNR.

USFD

- The Team developed the experimental paradigm 'Joystick Task' enabling the study of intrinsic motivations in humans and animals.
- With colleagues in Japan, the Team developed a variant of the 'Joystick Task' using monkeys and saccadic eye movements, allowing us to more directly study brain mechanisms involved with action acquisition and already producing a wealth of data (e.g., signals originating exclusively from subcortical visual processing is sufficient for development of novel actions).
- The Team completed experimental studies with humans showing that action acquisition using visual reinforcement signals that are not directly-available to subcortical structures is impaired, supporting neuroscientific theories underlying the IM-CLeVeR project.
- The Team developed a biologically-inspired neural network model of the 'Joystick Task' that describes 'intelligent' exploration strategies using only simple mechanisms of the basal ganglia. Several testable predictions are being tested with experimental studies.
- The Team developed a model showing how a period of stable behaviour can arise from a learning rule that does not incorporate any notion of optimality; such behaviour diverges with overtraining, providing a potential explanation for why dopamine signal must habituate in novelty-based action discovery.

FIAS

- The Team implemented on the iCub a bio-inspired general-purpose vision system capable of autonomously exploring the environment and determining the subset of relevant objects that will be subsequently learnt for future recognition based on intrinsic motivations.
- The Team developed new biologically plausible systems for vergence control and eye-head coordination learning based on intrinsic motivations.

UU

- The Team successfully developed a new approach for segmentation of objects using live streams from robot cameras in natural scenes using a 3D approach. Successfully implemented a Bayesian based method for hierarchical representation of data and demonstrated that the approach is an effective method for information storage for a hierarchy of images.
- The Team developed an online method of detecting features and recognising objects for the novelty detector. Developed a new novelty detector that addresses limitations of previous ones and was validated in experiments with physical robots in real world environments.
- The Team implemented techniques to enable a physical robot to carry out actions on perceived objects and identify the outcome of these actions so that basic affordances of the objects can be associated with particular events. In particular, developed an evolutionary approach for robots to create new skills based on an appropriate combinations and sequencing of lower lever skills.

AU

- The Team conducted an extensive literature review of infant psychological and neurological development from conception to 12 months postnatal, which is underpinning work on staged development in the iCub.
- The Team implemented a number of models for staged eye and head saccade learning, and constrained reaching, on the iCub.
- The Team implemented robotic systems whose development is based on different types of constraints,

that it contributed to classify, and intrinsic motivations.

- The Team designed and implemented several interfaces between computational models used by different partners in relation to CLEVE-B models.

IDSIA-SUPSI

- The Team developed a humanoid planning framework to create task-relevant roadmaps, which can be used to perform smooth motions and build a basis for learning task-relevant behaviours.
- The Team extended the capabilities of icVision, a computer vision and hand-eye coordination framework, which allows in combination with CGP-IP, our Cartesian Genetic Programming implementation, the learning of visual representations of objects in the scene. These allow detection, identification and localisation in real-time, which is an important requirement for achieving manipulation. In particular, thanks to a collaboration with FIAS, we can autonomously explore the scene and learn our representations without the need of a human.
- The Team developed MoBeE, a modular behavioral environment for humanoids and other robots, which integrates elements from vision, planning, and control, in order to facilitate the synthesis of autonomous, adaptive behaviors.
- The Team developed Modular-Least Squares Policy Iteration (M-LSPI), a novel method that enables real-world application of LSPI to massive Markovian reinforcement learning problems through modular/hierarchical decomposition.
- The Team created the Upper Confidence Weighted Learning (UCWL) framework for calculating intrinsic rewards through estimating the confidence intervals of the agent's predictions, which allows for efficient exploration in human-robot interaction scenarios with incomplete feedback.
- The Team introduced Curiosity-Driven Modular Incremental Slow Feature Analysis (CD-MISFA), a hierarchical curiosity-driven learning system that autonomously learns multiple abstract slow-feature based representations from a robot's high-dimensional visual input stream.
- The Team introduced PowerPlay, a way of automatically inventing the simplest still unsolvable problems. Conducted first successful PowerPlay experiments based on recurrent neural networks.
- Based on successful collaboration between the project partners, IDSIA and FIAS integrated various vision-based methods in the Clever-K demonstrator and published their results together.

Where the project converged to a consensus position

We list below a number of key issues on which the consortium achieved a consensus:

Experimental paradigms

- The joy-stick task is an exciting novel experimental paradigm for investigating intrinsically and extrinsically motivated action acquisition. The paradigm can be used with many species (e.g., rats, monkeys, human adults, children). The paradigm is very flexible, cheap and easy to use.
- The mechatronic board is an important tool that can be used to study intrinsic motivations. The mechatronic board is endowed with a number of features that make it ideally suited to investigate intrinsic motivations in primates (monkeys, human adults, children) and humanoid robots: (a) it allows recording of multimodal information for behavioural analysis; (b) it allows the generation of complex stimuli (perceptual, motor, curiosity,...) that can elicit intrinsic motivations; (c) it allows the synchronization of different mechatronic modules and units; (d) it is fully reconfigurable (scalable, modular) and programmable, so idea to generate novel situations and causal changes needed to probe intrinsic motivations.

Brain, intrinsic motivations, cumulative learning

- In animals, short-latency dopamine signals can be caused by both extrinsic and intrinsic learning signals. These learning signals, and the mechanisms that generate them, can be usefully transferred to control the architectures driving the autonomous learning of robots.

- In animals there exist different systems of intrinsic motivations. Those investigated in the project involve: (a) the system pivoting on the superior colliculus, involved in sensory prediction error; (b) the system pivoting on hippocampus, involved in novelty detection; (c) an “agency system”, for now identified only at the functional level.
- In animals, the detection of ‘agency’ is itself intrinsically motivating and drives the agent to further explore what is being done to cause the sensory event, and to refine the action that leads to it. This ‘agency bonus’ is also a feature that could be incorporated into biomimetic control architectures.
- The vertebrate basic basal ganglia-cortical loop component, and the more complex architecture formed by the motor, associative, and limbic loops, has a pivotal role in action-selection and decision-making in animals. The relation between such loops creates a hierarchical architecture giving rise to habitual and goal-directed behaviour, two forms of behaviour at the basis of cumulative learning. Such architecture can also be used in bio-constrained models to control the behaviour of embodied/robotic artificial agents.
- Goals generation, management, and exploitation is pivotal for intrinsically motivated cumulative learning in animals and, we expect, in artificial systems.

Machine learning, robotics

- Intrinsic motivations and hierarchical architectures are the two fundamental elements for building autonomous cumulative learning robots.
- Information theory is fundamental to suggest new ways to implement intrinsic motivations, e.g. in terms of measures of information compression, KL-divergence, mutual information, empowerment, etc.
- Reinforcement learning is the most important learning paradigm to implement cumulative learning based on intrinsic motivations. In particular, cumulative learning requires to develop innovative *hierarchical* reinforcement learning systems.
- The problem of transfer (e.g., in transfer reinforcement learning) is a critical problem to have an effective cumulative learning.
- Robots with multiple-degrees of freedom are necessary to solve tasks of interest and to undergo a real open-ended development. However, this generates a large motor space very difficult to be searched, e.g. by learning algorithms. The solution to this problem is to use abstract representations of such spaces, for example dynamic movement primitives or algorithms that work in the operational space rather than in the joint space while relying on effective ways to map between the two.
- A critical aspects for studying autonomous cumulative learning is the availability of robotic architectures that do not break when they engage in autonomous interactions with the environment. The iCub robot has very desirable features for developmental robotic studies (similarity with human body in terms of sensors and actuators) but on the other side is very delicate: this problem was partially solved with the use of a “virtual skin” to protect the robot from self-collisions and dangerous collisions with the environment. However, future research on cumulative learning might greatly benefit of the availability of robotic platforms that are robust enough to freely interact with the surrounding physical environment without breaking (e.g., during an initially quasi-random motor babbling).

Notes on the multiple distinct approaches that need to be investigated and reconciled in future work

We list below a number of key issues on which the consortium *did not* achieved a consensus, so further investigations and comparisons should address them in future work:

Topic	One position...	...and the other
Differences between intrinsic	CNR-LOCEN It is possible to distinguish between IM and EM on the basis of their typical mechanisms	CNR-Barto There is a continuum between extrinsic and intrinsic motivations: in particular, they are

motivations (IM) and extrinsic motivations (EM)	and functions. Of course reality (especially biological reality) “is dirty” and mixes the two in various ways. However, the distinction <i>is</i> possible and has a great heuristic power to understand brain and behaviour, and to build robots.	located along a continuum in terms of distances from the final fitness.
Knowledge-based IM (KB-IM) and competence-based IM (CB-IM)	CNR-LOCEN, CNR-Barto KB-IM and CB-IM are distinct: - KB-IM are based on measures of the level of the acquired knowledge (or rate of acquisition): knowledge has to do with the capacity to predict future stimuli (prediction-based IM) or the posses of representations of stimuli in memory (novelty-based IM) - CB-IM are based on measures of the level of the acquired competence (or rate of acquisition): competence has to do with the capacity to accomplish desired states (goals). Note: here we used the term “knowledge” in the meaning introduced by Oudeyer for KB-IM, but in general knowledge also includes competence.	IDSIA-SUPSI Intrinsic motivation is due to the learning progress of a machine that encodes ALL data encountered by the learning agent – not only its standard sensory perceptions, but also its actions, and the programs that encode its skills. This view subsumes and unifies competence-based IM and prediction-based IM and other types of IM.
The architectures needed for cumulative learning	CNR-LOCEN, CNR-Barto We need well crafted architectures to support cumulative learning. These architectures have to be based on key elements: - Representations of stimuli at different levels of abstraction - Skills - Goals - Inverse models - Forward models Such elements: - Have to be suitably combined within the architectures in order to have cumulative learning - Have to be trained based on different types of intrinsic motivations	IDSIA-SUPSI A way of implementing cumulative learning in practice is to look at a general problem solver architecture that in principle allows for encoding all kinds of skills, from prediction algorithms to motor programs. IM is then either about inventing new problems and learning new corresponding skills without forgetting old ones, or about refining previous skills, e.g., by speeding them up or compressing them.
Novelty-based vs. Prediction-based IM	CNR-ISTC, CNR-ISTC, UU They are distinct: - Novelty-based IM are related to the presence or not in memory of a representation of the incoming stimulus. Novelty signals/motivations are triggered when the incoming stimulus is not found memory. - Prediction-based IM are related to the prediction of the incoming stimulus.	IDSIA-SUPSI They are the same as both involve some form of information compression. Prediction is a form of compression. Memory is a form of compression.

	Prediction-based signals/motivations are triggered when there is a mismatch between the prediction and the incoming stimulus.	
Bio-constrained and bio-inspired modelling	CNR-LOCEN To have cumulative modeling of developmental phenomena we need to build models constrained: - at the level of the reproduced behaviour, on the basis of behavioural and developmental data - at the level of the architecture learning and functioning, on the basis of known anatomy and physiology of brain	AU To study development, it is enough to reproduce behavioral and developmental phenomena on the basis of systems that have a general biological plausibility (e.g., mappings between fields) but it is OK if they are not constrained in terms of neuroscience.
System-level modelling of brain	USFD We agree we have to have system-level models, and to embodied them, to understand the functioning of the specific investigated part of brain. However, it is OK to represent in detailed neural ways the specific investigated part of brain and to engineer or hack the rest around it.	CNR-LOCEN No, we should aim to have system-level models with all parts expressed in the same neural formalism as the function often emerges from the interaction of parts, i.e. from the whole system. Hacking/engineering the “peripheral parts” should be kept at a minimum, or at least closely checked, as it tends to cause false problems or false solutions in relation to the critical studied components. Technically, building whole systems leads to the problem of building complex models difficult to understand and to publish: this problem can however be ameliorated by standardizing the neural representations of the various parts of the model.